

Optimal Power Allocation for Channel Estimation of OFDM Uplinks in Time-Varying Channels

Rugui Yao, Yinsheng Liu, Geng Li, and Juan Xu

This paper deals with optimal power allocation for channel estimation of orthogonal frequency-division multiplexing uplinks in time-varying channels. In the existing literature, the estimation of time-varying channel response in an uplink environment can be accomplished by estimating the corresponding channel parameters. Accordingly, the optimal power allocation studied in the literature has been in terms of minimizing the mean square error of the channel estimation. However, the final goal for channel estimation is to enable the application of coherent detection, which usually means high spectral efficiency. Therefore, it is more meaningful to optimize the power allocation in terms of capacity. In this paper, we investigate capacity with imperfect channel estimation. By exploiting the derived capacity expression, an optimal power allocation strategy is developed. With this developed power allocation strategy, improved performance can be observed, as demonstrated by the numerical results.

Keywords: OFDM, capacity, time-varying channel, uplink channel, power allocation.

I. Introduction

Orthogonal frequency-division multiplexing (OFDM) modulation can transmit data in parallel by modulating a number of orthogonal subcarriers and has been widely used in modern communication systems [1]. In an OFDM system, the frequency-selective channel is converted into multiple flat fading subchannels, which can greatly simplify the design of the equalizer in the receiver [2].

The optimization problem for the channel estimation of an OFDM system has been widely addressed in existing literature [3]–[7]. For pilot-based channel estimation, an optimal interpolation can be achieved using a two-dimensional Wiener filter [3]. However, due to the complexity of its implementation, an optimal Wiener interpolator cannot be used in practice; thus, other interpolators have to be adopted [6]–[7]. The optimal pilot pattern in terms of sampling efficiency has been suggested in [4] and further addressed in [5]. The optimal pilot design for capacity maximization is also investigated [8]. However, since pilot-based channel estimation assumes a channel to be unchanged in one OFDM symbol, the corresponding optimization is not valid for the situation where a channel changes inside of one OFDM symbol.

For an even faster time-varying channel response, a series of basis expansion model (BEM)-based algorithms have been proposed in [9]–[12], where different kinds of basis functions are employed to model the time-varying channel. Mathematically, the BEM can be considered as an application of the rank reduction decomposition [13], and an optimal decomposition strategy has been developed in terms of minimizing the mean square error (MSE) [12]. Although BEM-based algorithms allow a channel to vary inside of one OFDM symbol, several drawbacks are obvious [14]. Due to

Manuscript received June 25, 2014; revised Sept. 3, 2014; accepted Sept. 11, 2014.

This work was supported in part by the National Natural Science Foundation of China (No. 61271416 and 61301093), the Aerospace support fund of China (No. 2013-HT-XGD), NPU Foundation for Fundamental Research (No. JCY20130132).

Rugui Yao (corresponding author, yaorg@nwpu.edu.cn) and Geng Li (785462603@qq.com) are with the School of Electronics and Information, Northwestern Polytechnical University, Xi'an, China.

Yinsheng Liu (09111035@bjtu.edu.cn) is with the Institute of Broadband Wireless Mobile Communications & the State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China.

Juan Xu (xuj@mail.nwpu.edu.cn) is with the School of Electronic and Control Engineering, Chang'an University, Xi'an, China.

the finite expansion order of basis functions for channel response, the inherent model error by truncation cannot be avoided. Moreover, the estimated model coefficients are only valid for one symbol duration; thus, they have to be re-estimated on a symbol-by-symbol basis.

Recently, algorithms for estimating time-varying channels in a macrocellular uplink environment have been proposed in [15]–[17]. Accordingly, the optimal power allocation ratio of training symbol to data symbol has been derived in [17] in terms of minimizing the MSE of the channel estimation. Although this particular kind of optimum power allocation strategy is useful for channel estimation, the final goal for channel estimation is to enable the application of coherent detection, which usually means high spectral efficiency. Therefore, the optimization should be more meaningful if carried out in terms of optimizing the capacity rather than MSE. As an extension of the research of [17], capacity is chosen as the performance indicator in this paper. Correspondingly, the power allocation strategy is re-derived in this paper, aiming to maximize the capacity for OFDM transmission with imperfect channel estimation.

The rest of this paper is organized as follows. The system model is described in Section II. The expression of capacity with imperfect channel estimation is derived in Section III. The practical capacity with the inverse channel detection (ICD) estimator is derived in Section IV. In Section V, the optimal power allocation strategy is discussed. Numerical results are presented in Section VI. Finally, the conclusion is given in Section VII.

II. System Model

1. Signal Model

For OFDM transmission in a time-varying channel, the detected signal can be represented in vector form as

$$\mathbf{y} = \mathbf{H}\mathbf{d} + \mathbf{w}, \quad (1)$$

where $\mathbf{d} = (d_0, d_1, \dots, d_{N-1})^T$ denotes the frequency domain symbols with covariance matrix $E(\mathbf{d}\mathbf{d}^H) = \sigma_d^2 \mathbf{I}$, and $\mathbf{w} = (w_0, w_1, \dots, w_{N-1})^T$ is the additive noise with covariance matrix $E(\mathbf{w}\mathbf{w}^H) = \sigma_w^2 \mathbf{I}$. Here, \mathbf{I} is an $N \times N$ identity matrix and N is the subcarrier number during the OFDM transmission. Usually, it is assumed that \mathbf{d} and \mathbf{w} have zero means. The symbol \mathbf{H} represents the channel matrix, whose (m, n) th element is given by

$$[\mathbf{H}]_{(m,n)} = \frac{1}{N} \sum_{l=0}^{L-1} \sum_{k=0}^{N-1} h[l, k] e^{j\frac{2\pi(n-m)k}{N}} e^{-j\frac{2\pi nl}{N}}, \quad (2)$$

where $h[l, k]$ denotes the time-varying channel response for the

current OFDM symbol. The variable k denotes the discrete sample time, and l denotes the index of the resolvable path. As shown in [17], $h[l, k]$ can be explicitly represented as

$$h[l, k] = h_l e^{j\frac{2\pi v_l k}{N}} \quad (3)$$

for the uplink environment, where h_l and v_l are the complex amplitudes and Doppler shifts of the l th path, respectively. This channel model is particularly suitable for the uplink transmission, where the angular spread is small for non-resolvable components of a channel tap. Therefore, the Doppler shifts for those non-resolvable components inside a channel tap are approximately the same and can thus be combined [18]. Similar models are also adopted in [19]–[20].

It should be noted that due to the fast time variation scenario, the channel response cannot hold unchanged during one OFDM transmission; thus, \mathbf{H} is not a diagonal matrix. This is essentially different from the slow time variation scenario, where the channel matrix is diagonal [21].

2. Channel Estimation

To recover the transmitted data symbols, the channel matrix \mathbf{H} should be known to the receiver. With the method proposed in [17], $h[l, k]$ can be regenerated by estimating the complex amplitudes and Doppler shifts. Let $\hat{h}[l, k]$ and $\hat{\mathbf{H}}$ denote the regenerated channel response and regeneration of \mathbf{H} , respectively. We then have

$$[\hat{\mathbf{H}}]_{(m,n)} = \frac{1}{N} \sum_{l=0}^{L-1} \sum_{k=0}^{N-1} \hat{h}[l, k] e^{j\frac{2\pi(n-m)k}{N}} e^{-j\frac{2\pi nl}{N}}. \quad (4)$$

Correspondingly, the estimation error for $h[l, k]$ and \mathbf{H} can be given, respectively, as $\tilde{h}[l, k] = \hat{h}[l, k] - h[l, k]$ and $\tilde{\mathbf{H}} = \hat{\mathbf{H}} - \mathbf{H}$. Likewise, a similar relation between $\tilde{h}[l, k]$ and $\tilde{\mathbf{H}}$ can be given as

$$[\tilde{\mathbf{H}}]_{(m,n)} = \frac{1}{N} \sum_{l=0}^{L-1} \sum_{k=0}^{N-1} \tilde{h}[l, k] e^{j\frac{2\pi(n-m)k}{N}} e^{-j\frac{2\pi nl}{N}}. \quad (5)$$

With the estimated channel matrix $\hat{\mathbf{H}}$, the transmitted data symbols can be recovered. However, due to the presence of the estimation error, the performance obtained is worse than that obtained with perfect channel estimation. Therefore, it is necessary to evaluate the system performance with imperfect channel estimation.

III. Capacity

In this paper, the capacity is chosen as the performance indicator since it is a good metric of the channel efficiency without having to consider implementation details.

1. Capacity with Perfect Channel Estimation

Assuming that the channel estimation is perfect, the capacity for OFDM transmission can be given as

$$C_{\text{perf}} = \frac{1}{N} \max_{p_d(\mathbf{d})} \{I(\mathbf{y}; \mathbf{d})\} \text{ bits/symbol}, \quad (6)$$

where $p_d(\cdot)$ denotes the probability density function of \mathbf{d} . The mutual information between \mathbf{y} and \mathbf{d} is denoted by $I(\mathbf{y}; \mathbf{d}) = H(\mathbf{y}) - H(\mathbf{y} | \mathbf{d})$, where $H(\cdot)$ denotes the information entropy. Similar to the derivation for the multi-input and multi-output (MIMO) system in [21], the mutual information in (6) can achieve a maximum only when \mathbf{d} is a Gaussian distributed random vector. In this case, the capacity for OFDM transmission with perfect channel estimation is given as [22]

$$C_{\text{perf}} = \frac{1}{N} \log_2 \det \left(\mathbf{I} + \frac{\sigma_d^2}{\sigma_w^2} \mathbf{H} \mathbf{H}^H \right), \quad (7)$$

where $\det(\cdot)$ denotes the determinant of a matrix.

Note that the capacity representations for an OFDM system and MIMO system are actually the same. This is not surprising because both systems have the same system model in (1). However, the difference is that for a MIMO system, the elements in \mathbf{H} can be considered as independently distributed, while for an OFDM system, the elements in \mathbf{H} are actually correlated.

2. Capacity with Practical Channel Estimation

Consider now that the estimate of \mathbf{H} is imperfect with an estimation error $\tilde{\mathbf{H}} = \hat{\mathbf{H}} - \mathbf{H}$. Since $\hat{\mathbf{H}}$ is known, the received signal in (1) can be rewritten as

$$\mathbf{y} = \hat{\mathbf{H}} \mathbf{d} + \mathbf{u}, \quad (8)$$

where $\mathbf{u} = \mathbf{w} - \tilde{\mathbf{H}} \mathbf{d}$ can be considered as the additive noise. Bearing in mind that \mathbf{d} has zero mean, the correlation matrix $\mathbf{R}_{\text{uu}} = E(\mathbf{u} \mathbf{u}^H)$ can be given as

$$\mathbf{R}_{\text{uu}} = \sigma_w^2 \mathbf{I} + \sigma_d^2 E(\tilde{\mathbf{H}} \tilde{\mathbf{H}}^H). \quad (9)$$

From (9), it is observed that the data symbol also contributes to the overall noise power due to the presence of channel estimation error.

With $\hat{\mathbf{H}}$ known to the receiver, the capacity for OFDM transmission in the presence of channel estimation error, as shown in (8), can be given as

$$C_{\text{prac}} = \frac{1}{N} \max_{p_d(\mathbf{d})} \{I(\mathbf{y}; \mathbf{d})\} \text{ bits/symbol}, \quad (10)$$

where

$$I(\mathbf{y}; \mathbf{d}) = H(\mathbf{d}) - H(\mathbf{d} | \mathbf{y}) \quad (11)$$

denotes the mutual information for the practical channel estimator.

To achieve the maximal mutual information, assume that the data vector \mathbf{d} is Gaussian distributed; thus, we have [21]

$$H(\mathbf{d}) = \log_2 \det(\pi e \sigma_d^2 \mathbf{I}). \quad (12)$$

Note that under this assumption, \mathbf{y} and \mathbf{u} are also Gaussian distributed since both vectors are a linear combination of Gaussian vectors.

Recalling the relation in (8), we find that the transmitted data symbol vector \mathbf{d} can be estimated linearly if given \mathbf{y} and $\hat{\mathbf{H}}$. Different kinds of linear receivers can be employed for this purpose. In this section, we adopt the minimum-MSE (MMSE) receiver [13]. As shown in [22], the MMSE receiver is information-lossless; thus, the following equation can hold

$$I(\mathbf{d}; \mathbf{y}) = I(\mathbf{d}; \hat{\mathbf{d}}_{\text{MMSE}}), \quad (13)$$

where $\hat{\mathbf{d}}_{\text{MMSE}}$ denotes the MMSE estimate of \mathbf{d} . From (13), it is easy to obtain $H(\mathbf{d} | \mathbf{y}) = H(\mathbf{d} | \hat{\mathbf{d}}_{\text{MMSE}})$, the right-hand side of which can be further rewritten as $H(\mathbf{d} | \hat{\mathbf{d}}_{\text{MMSE}}) = H(\tilde{\mathbf{d}}_{\text{MMSE}})$, where $\tilde{\mathbf{d}}_{\text{MMSE}} = \hat{\mathbf{d}}_{\text{MMSE}} - \mathbf{d}$ denotes the estimation error for the MMSE estimator. Therefore, we can deduce that

$$H(\mathbf{d} | \mathbf{y}) = H(\tilde{\mathbf{d}}_{\text{MMSE}}). \quad (14)$$

Noting that \mathbf{y} is a Gaussian distributed vector, $\hat{\mathbf{d}}_{\text{MMSE}}$ is therefore also Gaussian distributed since it is a linear composition of \mathbf{y} . Correspondingly, $\tilde{\mathbf{d}}_{\text{MMSE}}$ is a Gaussian vector as well; thus, we have

$$\begin{aligned} H(\mathbf{d} | \mathbf{y}) &= H(\tilde{\mathbf{d}}_{\text{MMSE}}) \\ &= \log_2 \det \left[\pi e E(\tilde{\mathbf{d}}_{\text{MMSE}} \tilde{\mathbf{d}}_{\text{MMSE}}^H) \right]. \end{aligned} \quad (15)$$

For an MMSE estimator, the estimation of \mathbf{d} can be represented as [13]

$$\hat{\mathbf{d}}_{\text{MMSE}} = \mathbf{W}^H \mathbf{y}, \quad (16)$$

where

$$\mathbf{W} = \left(\hat{\mathbf{H}} \hat{\mathbf{H}}^H + \frac{\mathbf{R}_{\text{uu}}}{\sigma_d^2} \right)^{-1} \hat{\mathbf{H}}. \quad (17)$$

Therefore, the correlation matrix of estimation error $\tilde{\mathbf{d}}_{\text{MMSE}}$ can be given as

$$\begin{aligned} &E(\tilde{\mathbf{d}}_{\text{MMSE}} \tilde{\mathbf{d}}_{\text{MMSE}}^H) \\ &= E \left[(\mathbf{W}^H \mathbf{y} - \mathbf{d})(\mathbf{W}^H \mathbf{y} - \mathbf{d})^H \right] \\ &= \sigma_d^2 \mathbf{I} - \sigma_d^2 \hat{\mathbf{H}}^H \left(\hat{\mathbf{H}} \hat{\mathbf{H}}^H + \frac{\mathbf{R}_{\text{uu}}}{\sigma_d^2} \right)^{-1} \hat{\mathbf{H}}; \end{aligned} \quad (18)$$

thus, the capacity in (10) can be derived as

$$C_{\text{prac}} = \frac{1}{N} \log_2 \det \left[\mathbf{I} - \hat{\mathbf{H}}^H \left(\hat{\mathbf{H}} \hat{\mathbf{H}}^H + \frac{\mathbf{R}_{\text{uu}}}{\sigma_d^2} \right)^{-1} \hat{\mathbf{H}} \right] \quad (19)$$

$$= \frac{1}{N} \log_2 \det \left(\mathbf{I} + \sigma_d^2 \hat{\mathbf{H}} \hat{\mathbf{H}}^H \mathbf{R}_{\text{uu}}^{-1} \right),$$

where the second equation is due to the matrix inversion lemma. Note that the derived capacity expression is similar to the representation derived in [23] for a MIMO system. This is not surprising because the OFDM transmission model is actually identical to the MIMO model.

From Appendix A, we find that $E(\tilde{\mathbf{H}}\tilde{\mathbf{H}}^H)$ is actually a diagonal matrix, assuming N is large enough, which can be represented as

$$E(\tilde{\mathbf{H}}\tilde{\mathbf{H}}^H) = \frac{L}{N} \left(\frac{\sigma_w^2}{\sigma_d^2} + \frac{2\sigma_w^2}{3\sigma_t^2} \right) \mathbf{I}, \quad (20)$$

where σ_t^2 is the power of the training symbol used in [17]. By substituting (20) into (19), the capacity for a practical channel estimator can be obtained as

$$C_{\text{prac}} = \frac{1}{N} \log_2 \det \left(\mathbf{I} + \varepsilon \hat{\mathbf{H}} \hat{\mathbf{H}}^H \right), \quad (21)$$

where

$$\varepsilon = \left[\frac{(N+L)\sigma_w^2}{N\sigma_d^2} + \frac{2L\sigma_w^2}{3N\sigma_t^2} \right]^{-1}. \quad (22)$$

IV. Extension to ICD Estimator

In this section, the practical capacity for OFDM transmission is derived considering the ICD estimator. Due to the information-loss characteristics of the ICD estimator [22], its achieving capacity will be degraded; that is,

$$I(\mathbf{d}; \mathbf{y}) \geq I(\mathbf{d}; \hat{\mathbf{d}}_{\text{ICD}}), \quad (23)$$

referring to (13). For the detected symbol $\hat{\mathbf{d}}_{\text{ICD}}$ with ICD estimator, the mutual information in (11) can be further computed as

$$I(\mathbf{d}; \hat{\mathbf{d}}_{\text{ICD}}) = H(\mathbf{d}) - H(\mathbf{d} | \hat{\mathbf{d}}_{\text{ICD}}) \quad (24)$$

$$= H(\mathbf{d}) - H(\tilde{\mathbf{d}}_{\text{ICD}}).$$

For an ICD estimator, the detected matrix, \mathbf{W}_{ICD} , can be represented as [22]

$$\mathbf{W}_{\text{ICD}} = \hat{\mathbf{H}} \left(\hat{\mathbf{H}}^H \hat{\mathbf{H}} \right)^{-1}. \quad (25)$$

And, the correlation matrix of estimation error $\tilde{\mathbf{d}}_{\text{ICD}}$ can be given as

$$E(\tilde{\mathbf{d}}_{\text{ICD}} \tilde{\mathbf{d}}_{\text{ICD}}^H) = E \left[(\mathbf{W}_{\text{ICD}}^H \mathbf{y} - \mathbf{d}) (\mathbf{W}_{\text{ICD}}^H \mathbf{y} - \mathbf{d})^H \right] \quad (26)$$

$$= \mathbf{R}_{\text{uu}} \left(\hat{\mathbf{H}}^H \hat{\mathbf{H}} \right)^{-1}.$$

Like (15), the information entropy, $H(\tilde{\mathbf{d}}_{\text{ICD}})$, in (24) can be computed from (26) as

$$H(\tilde{\mathbf{d}}_{\text{ICD}}) = \log_2 \det \left[\pi e E(\tilde{\mathbf{d}}_{\text{ICD}} \tilde{\mathbf{d}}_{\text{ICD}}^H) \right] \quad (27)$$

$$= \log_2 \det \left[\pi e \mathbf{R}_{\text{uu}} \left(\hat{\mathbf{H}}^H \hat{\mathbf{H}} \right)^{-1} \right].$$

Substituting both (12) and (27) in (24), the practical capacity with ICD estimator can be derived as

$$C_{\text{prac}}^{\text{ICD}} = \frac{1}{N} \log_2 \det \left(\sigma_d^2 \hat{\mathbf{H}}^H \hat{\mathbf{H}} \mathbf{R}_{\text{uu}}^{-1} \right). \quad (28)$$

V. Optimal Power Allocation

To maximize the capacity, we just consider the capacity with MMSE estimator in this section.

The optimal power allocation aims to maximize the channel capacity, subject to

$$\sigma_t^2 + \sigma_d^2 = P, \quad (29)$$

where P is the total power. A similar problem has been addressed in [17], where the optimization aims to minimize the MSE of the channel estimation. In this paper, our purpose is to maximize the capacity with given P . Considering the random fading in a wireless channel, the capacity in (10) is not directly used. Instead, we adopt the average capacity as the objective function, which can be defined as

$$\bar{C}_{\text{prac}} = E \left[\frac{1}{N} \log_2 \det \left(\mathbf{I} + \varepsilon \hat{\mathbf{H}} \hat{\mathbf{H}}^H \right) \right]. \quad (30)$$

To find the optimal power allocation ratio, which can maximize the average capacity above, we define $\sigma_t^2 = P\alpha$; thus, $\sigma_d^2 = P(1-\alpha)$, where $\alpha \in (0, 1)$ is the power allocation ratio. Correspondingly, ε is also a function of α , that is, $\varepsilon = \varepsilon(\alpha)$ and $\bar{C}_{\text{prac}} = \bar{C}_{\text{prac}}(\alpha)$. Rewriting ε as a function of the ratio α , we have

$$\varepsilon(\alpha) = \frac{P}{\sigma_w^2} \cdot \frac{3N\alpha(1-\alpha)}{(3N+L)\alpha + 2L}. \quad (31)$$

Bearing in mind that $\hat{\mathbf{H}} \hat{\mathbf{H}}^H$ is a Hermitian matrix, the eigenvalue decomposition of $\hat{\mathbf{H}} \hat{\mathbf{H}}^H$ can be expressed as

$$\hat{\mathbf{H}} \hat{\mathbf{H}}^H = \mathbf{P} \mathbf{A} \mathbf{P}^H, \quad (32)$$

where \mathbf{P} is a unitary matrix, and $\mathbf{A} = \text{diag}(\lambda_0, \lambda_1, \dots, \lambda_{N-1})$ where λ_n is the n th eigenvalue. For a Hermitian matrix, λ_n is

real for any $n \in (0, N-1)$, and the eigenvalues have the following relationship: $\lambda_0 \geq \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$. With the relation in (32), the average capacity can be derived as

$$\begin{aligned} \bar{C}_{\text{prac}} &= E \left\{ \frac{1}{N} \log_2 \det \{ \mathbf{P} [\mathbf{I} + \varepsilon(\alpha) \mathbf{A}] \mathbf{P}^H \} \right\} \\ &= E \left\{ \frac{1}{N} \log_2 \det [\mathbf{I} + \varepsilon(\alpha) \mathbf{A}] \right\} \\ &= \frac{1}{N} \sum_{n=0}^{N-1} E \{ \log_2 [1 + \varepsilon(\alpha) \lambda_n] \}. \end{aligned} \quad (33)$$

Since $\log_2(x)$ is an increasing function of x , an upper bound for (33) can be given as

$$\bar{C}_{\text{prac}} \leq E \{ \log_2 [1 + \varepsilon(\alpha) \lambda_{\max}] \} \leq \log_2 [1 + \varepsilon(\alpha) \bar{\lambda}_{\max}] \quad (34)$$

if denoting λ_{\max} as the maximal eigenvalue, and $\bar{\lambda}_{\max} = E(\lambda_{\max})$ denotes the mean of λ_{\max} . The second inequality in (34) is due to Jensen's inequality [21]. In Appendix B, it is shown that

$$\bar{\lambda}_{\max} = L\sigma_h^2 + \frac{L}{N} \left(\frac{\sigma_w^2}{\sigma_d^2} + \frac{2\sigma_w^2}{3\sigma_t^2} \right), \quad (35)$$

where

$$\sigma_h^2 = \frac{1}{L} \sum_{l=0}^{L-1} E(|h_l|^2). \quad (36)$$

For an individual path, $E(|h_l|^2)$ depends on the channel delay profile. Note that $\bar{\lambda}_{\max}$ is also a function of α , that is,

$$\bar{\lambda}_{\max}(\alpha) = L\sigma_h^2 + \left(\frac{L\sigma_w^2}{NP} \cdot \frac{\alpha + 2}{3\alpha(1-\alpha)} \right). \quad (37)$$

Therefore, the upper bound for \bar{C}_{prac} can be written as

$$\bar{C}_{\text{prac}}^U(\alpha) = \log_2 [1 + \varepsilon(\alpha) \bar{\lambda}_{\max}(\alpha)]. \quad (38)$$

By solving the equation

$$\frac{d\bar{C}_{\text{prac}}^U(\alpha)}{d\alpha} = 0, \quad (39)$$

the optimal power allocation ratio, α_{opt} , for maximizing the capacity upper bound can be obtained. As it will be shown, \bar{C}_{prac} can achieve the maximum at α_{opt} as well.

Substituting (31) and (37) into (39), (39) can be rewritten as

$$\begin{aligned} & \left[\frac{P}{\sigma_w^2} \cdot \frac{-3N\alpha^2 + 3N\alpha}{(3N+L)\alpha + 2L} \right] \frac{d \left[\frac{L\sigma_w^2}{NP} \cdot \frac{\alpha + 2}{-3\alpha^2 + 3\alpha} \right]}{d\alpha} \\ & + \frac{d \left[\frac{P}{\sigma_w^2} \cdot \frac{-3N\alpha^2 + 3N\alpha}{(3N+L)\alpha + 2L} \right]}{d\alpha} \left[\frac{L\sigma_w^2}{NP} \cdot \frac{\alpha + 2}{-3\alpha^2 + 3\alpha} \right] = 0, \end{aligned} \quad (40)$$

which after some algebraic manipulation, can be simplified as

$$a\alpha^2 + b\alpha + c = 0, \quad (41)$$

where $a = 3LNP\sigma_h^2(L+3N)$, $b = 12L^2NP\sigma_h^2$, and $c = 6LN(\sigma_w^2 - LP\sigma_h^2)$. Equation (41) can be easily solved using the well-known root formula, and the solution is given as

$$\alpha_{\text{opt}} = \frac{-b + \sqrt{b^2 - 4ac}}{2a}, \quad (42)$$

where α_{opt} is a positive value.

VI. Numerical Results

Numerical results are shown in this section to demonstrate the efficiency of the derived power allocation ratio. An OFDM transmission consisting of $N = 64$ subcarriers is investigated in the numerical analysis. For the wireless channel, $L = 3$ paths are considered, with power delay profile as $E(|h_l|^2) = 1$ for $l = 0, 1, 2$ and Doppler shifts as 0.2, 0.1, and 0.1. The algorithm proposed in [17] is adopted for estimating the channel response. The SNR is defined as $\text{SNR} = P / \sigma_w^2$.

The numerical results and theoretical upper bounds are shown in Fig. 1. It is observed that the capacity, as well as the upper bounds, are convex functions of α . Both the numerical results and the upper bounds can achieve their respective maximums at $\alpha_{\text{opt}} \approx 0.15$, which coincides with our theoretical prediction with (42) for $N = 64, L = 3, \sigma_h^2 = 1$, and $P = 1$.

The capacities for different power allocation ratios are presented in Fig. 2. For $\alpha = 0.15$, the theoretical upper bound

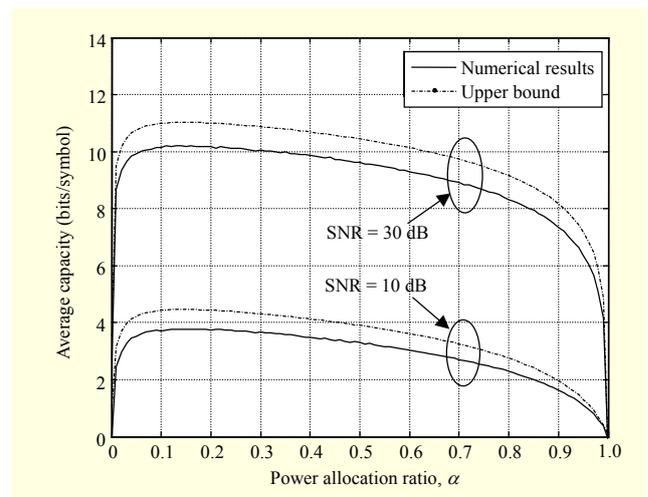


Fig. 1. Capacity vs. power allocation ratio for SNR = 10 dB and SNR = 30 dB, respectively. Theoretical upper bounds are also presented for comparison.

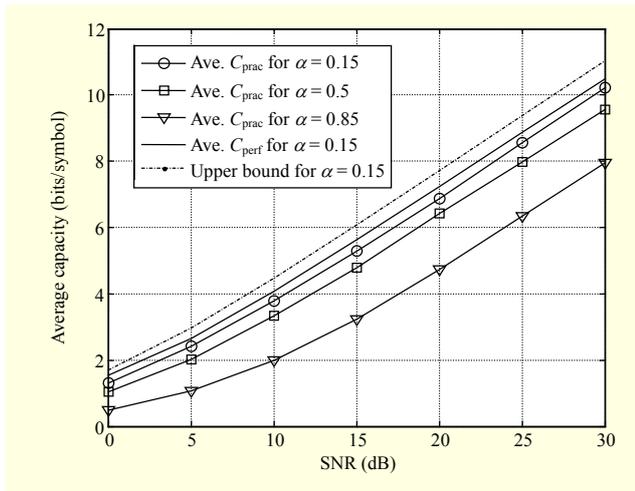


Fig. 2. Average capacities for $\alpha = 0.15, 0.5, 0.85$, respectively in terms of SNR. The upper bound and average capacity with perfect channel estimation are also presented.

and capacity with perfect channel estimation (that is, (7)) are also shown for comparison. As it can be seen from Fig. 2, the performances for different α 's are significantly different. For $\alpha = 0.85$, the worst performance can be observed. On the other hand, a better performance can be obtained by adopting $\alpha = 0.5$. This can be explained by referring to Fig. 1. As seen in Fig. 1, $\alpha = 0.85$ is relatively far from the arrest point ($\alpha \approx 0.15$) compared to $\alpha = 0.5$. This accounts for the different performances we observed since \bar{C}_{prac} is a convex function of α . Also, from Fig. 2, we can observe a 1 dB performance gap between \bar{C}_{prac} and \bar{C}_{perf} when $\alpha = 0.15$. This suggests that the ideal capacity with perfect channel estimation can be almost achieved by adopting the derived optimal power allocation ratio. Meanwhile, a 2.5 dB performance gap between \bar{C}_{prac} with $\alpha = 0.15$ and the theoretical upper bound can be observed due to the inequality in (34).

In Fig. 3, the cumulative distribution functions (CDFs) of the capacities are also shown. Three cases, where α takes 0.15, 0.5, and 0.85, are investigated in Fig. 3. Similar to the situation in Fig. 2, an obvious performance improvement can be observed by adopting the optimal power allocation ratio, and the performance gets worse as α gets further away from the arrest point. Also, we can find that the probability of achieving a capacity of more than 2.5 bits/symbol is 0.9 at SNR = 10 dB. At SNR = 30 dB, the capacity can achieve more than 8.5 bits/symbol at a probability of 0.9. For both situations, the capacity achieved using optimal power allocation is very close to the capacity with perfect channel estimation, demonstrating the efficiency of the proposed power allocation strategy.

The comparison of average capacities for the proposed algorithms in [15]–[17] is shown in Fig. 4. The average capacity

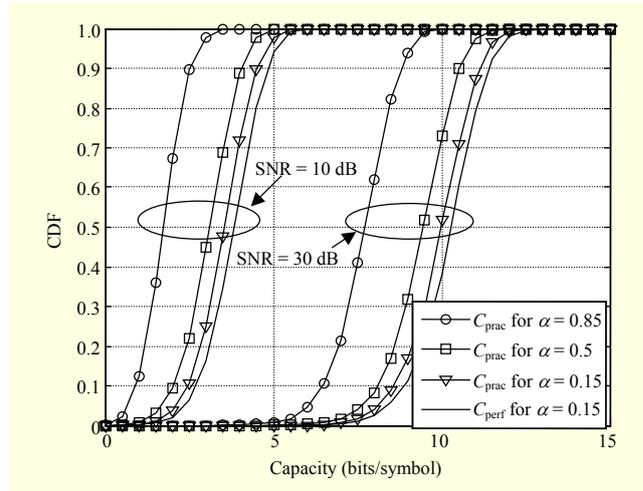


Fig. 3. CDFs for SNR = 10 dB and SNR = 30 dB, respectively. CDF of capacity with perfect channel estimation are also presented.

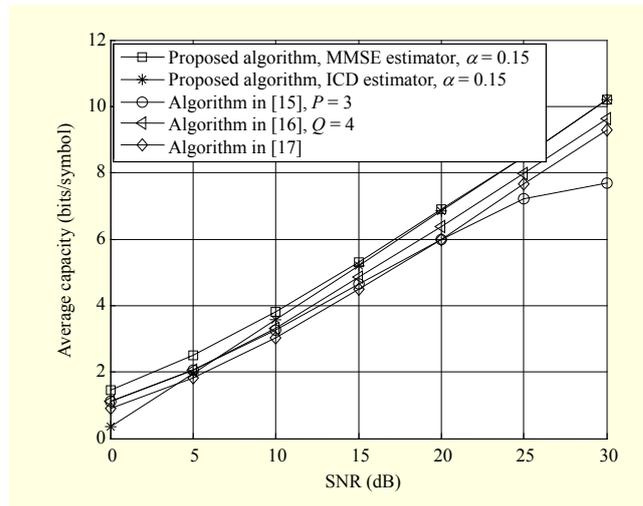


Fig. 4. Comparison of average capacities for the proposed algorithm and algorithms in [15]–[17]. Orders of the Taylor expansions in [15] and [16] are represented by P and Q , respectively.

of the proposed algorithm with ICD estimator is also presented for comparison. As shown in Fig. 4, due to the noise enhancement issue and information loss characteristics, the ICD estimator leads to a worse performance at lower SNR, compared with the MMSE estimator. With the SNR increasing, the capacity with ICD estimator approached the capacity of the MMSE estimator. Note that the training symbol and data symbols are with equal power allocation for the algorithms in [15] and [16]. For the algorithm in [17], the optimal power allocation is based on minimizing the MSE of the channel estimation. As it can be seen from Fig. 4, our proposed algorithm outperforms all the algorithms in [15]–[17]. Even at high SNR, there exists a larger gap between the proposed

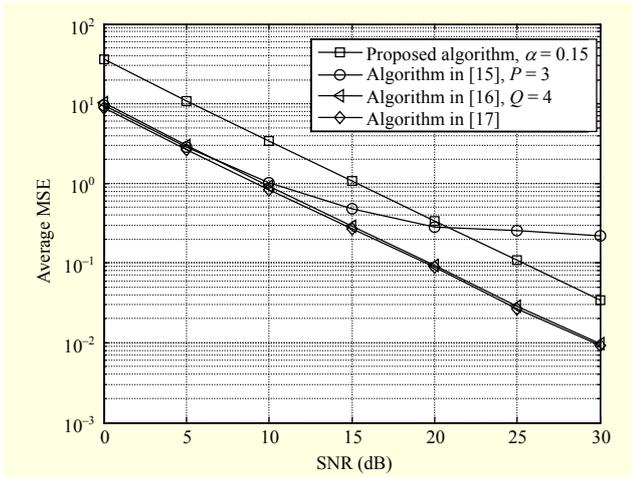


Fig. 5. Average MSEs for proposed algorithm and algorithms in [15]–[17] at different SNRs.

algorithm and that in [15], which can be interpreted later by the average MSE results in Fig. 5.

The average MSEs for the different algorithms in [15]–[17] and this paper are compared in Fig. 5. Because less power was allocated for the training symbol, our proposed algorithm presents worse MSE performance; however, it achieves the best capacity, as shown in Fig. 4. To minimize the MSE of the channel estimation, the algorithm in [17] allocates more power to the training symbol. Nevertheless, this scheme is not efficient in terms of capacity performance, as can be seen from Fig. 4. The algorithm in [16] has almost the same MSE as the algorithm in [17]. The algorithm in [15] obtains as superior an MSE performance as those in [16] and [17] at low SNR; however, at high SNR, it does not achieve good MSE performance and further contributes to the poor capacity performance, as shown in Fig. 4. This results from the large intrinsic estimation error of the algorithm in [15].

VII. Conclusion

In this paper, the optimal power allocation for channel estimation of OFDM uplinks in time-varying channels has been investigated. As an extension of the research of [17], this paper intends to derive the optimal power allocation ratio in terms of maximizing the capacity, instead of minimizing the MSE of the channel estimation, as it has been done in [17]. Based on the derived upper bound expression for the capacity with imperfect channel estimation, the optimal power allocation strategy can be easily obtained by reasonably allocating the total power between the training symbol and data symbols. The numerical results are also shown, demonstrating the efficiency of the derived power allocation strategy.

Appendix A: Derivation of (20)

Recalling the representation of $\tilde{\mathbf{H}}$ in (5), it actually can be written as

$$[\tilde{\mathbf{H}}]_{(m,n)} = \frac{1}{N} \sum_{k=0}^{N-1} e^{-j\frac{2\pi mk}{N}} \tilde{\mathbf{h}}_k^T \mathbf{q}_k(n), \quad (\text{A.1})$$

where

$$\tilde{\mathbf{h}}_k = (\tilde{h}[0, k], \tilde{h}[1, k], \dots, \tilde{h}[L-1, k])^T \quad (\text{A.2})$$

and

$$\mathbf{q}_k(n) = \left(e^{j\frac{2\pi kn}{N}}, e^{j\frac{2\pi(k-1)n}{N}}, \dots, e^{j\frac{2\pi(k-L+1)n}{N}} \right)^T. \quad (\text{A.3})$$

Define $\tilde{\mathbf{H}}^T = (\tilde{\mathbf{H}}_0, \tilde{\mathbf{H}}_1, \dots, \tilde{\mathbf{H}}_{N-1})$, where $\tilde{\mathbf{H}}_m$ denotes the m th column of matrix $\tilde{\mathbf{H}}^T$. Then, $\tilde{\mathbf{H}}_m^T$ can be represented as

$$\tilde{\mathbf{H}}_m^T = \frac{1}{N} \sum_{k=0}^{N-1} e^{-j\frac{2\pi mk}{N}} \tilde{\mathbf{h}}_k^T \mathbf{Q}_k, \quad (\text{A.4})$$

where

$$\mathbf{Q}_k = [\mathbf{q}_k(0), \mathbf{q}_k(1), \dots, \mathbf{q}_k(N-1)]. \quad (\text{A.5})$$

With the definition of $\tilde{\mathbf{H}}$, the (m, n) th element of $E(\tilde{\mathbf{H}}\tilde{\mathbf{H}}^H)$ can be given as

$$E(\tilde{\mathbf{H}}_m^T \tilde{\mathbf{H}}_n^*) = \frac{1}{N^2} \sum_{k=0}^{N-1} \sum_{t=0}^{N-1} e^{j\frac{2\pi n(t-mk)}{N}} E(\tilde{\mathbf{h}}_k^T \mathbf{M}_{k,t} \tilde{\mathbf{h}}_t^*), \quad (\text{A.6})$$

where $\mathbf{M}_{k,t} = \mathbf{Q}_k \mathbf{Q}_t^H$. The term $E(\tilde{\mathbf{h}}_k^T \mathbf{M}_{k,t} \tilde{\mathbf{h}}_t^*)$ in (A.6) can be rewritten as

$$E(\tilde{\mathbf{h}}_k^T \mathbf{M}_{k,t} \tilde{\mathbf{h}}_t^*) = \sum_{l_1=0}^{L-1} \sum_{l_2=0}^{L-1} E(\tilde{h}[l_1, k] \tilde{h}^*[l_2, t]) [\mathbf{M}_{k,t}]_{(l_1, l_2)}, \quad (\text{A.7})$$

where $[\mathbf{M}_{k,t}]_{(l_1, l_2)}$ denotes the (l_1, l_2) th element of $\mathbf{M}_{k,t}$.

From the derivation in [17], we can conclude that the estimation errors for different paths are independent; that is,

$$E(\tilde{h}[l_1, k] \tilde{h}^*[l_2, t]) \begin{cases} \neq 0 & \text{for } l_1 = l_2, \\ = 0 & \text{for } l_1 \neq l_2. \end{cases} \quad (\text{A.8})$$

Therefore, the equation in (A.7) can be simplified as

$$\begin{aligned} E(\tilde{\mathbf{h}}_k^T \mathbf{M}_{k,t} \tilde{\mathbf{h}}_t^*) &= \sum_{l=0}^{L-1} E(\tilde{h}[l, k] \tilde{h}^*[l, t]) [\mathbf{M}_{k,t}]_{(l,l)} \\ &= E(\tilde{\mathbf{h}}_k^T \text{diag}\{\mathbf{M}_{k,t}\} \tilde{\mathbf{h}}_t^*), \end{aligned} \quad (\text{A.9})$$

where $\text{diag}\{\mathbf{M}_{k,t}\}$ denotes a diagonal matrix composed of the elements on the diagonal line of $\mathbf{M}_{k,t}$. Recalling that $\mathbf{M}_{k,t} = \mathbf{Q}_k \mathbf{Q}_t^H$ and the definition of \mathbf{Q}_k in (A.5), we can obtain

that

$$\begin{aligned} \text{diag}\{\mathbf{M}_{k,t}\} &= \sum_{n=0}^{N-1} \text{diag}\{\mathbf{q}_k(n)\mathbf{q}_t^H(n)\} \\ &= \sum_{n=0}^{N-1} \begin{pmatrix} e^{\frac{j2\pi(k-t)n}{N}} & & \\ & \ddots & \\ & & e^{\frac{j2\pi(k-t)n}{N}} \end{pmatrix} \end{aligned} \quad (\text{A.10})$$

From (A.10), we can observe that $\text{diag}\{\mathbf{M}_{k,t}\}$ is a zero matrix when $k \neq t$ and that $\text{diag}\{\mathbf{M}_{k,t}\} = N \cdot \mathbf{I}$ when $k = t$.

Therefore, the relation in (A.9) can be written as

$$\mathbb{E}(\tilde{\mathbf{h}}_k^T \mathbf{M}_{k,t} \tilde{\mathbf{h}}_t^*) = \begin{cases} N \cdot \mathbb{E}(\tilde{\mathbf{h}}_k^T \tilde{\mathbf{h}}_k^*) & \text{for } k = t, \\ 0 & \text{for } k \neq t. \end{cases} \quad (\text{A.11})$$

With the relation in (A.11), the (m, n) th element of $\mathbb{E}(\tilde{\mathbf{H}}\tilde{\mathbf{H}}^H)$ in (A.6) can be simplified as

$$\begin{aligned} \mathbb{E}(\tilde{\mathbf{H}}_m^T \tilde{\mathbf{H}}_n^*) &= \frac{1}{N} \sum_{k=0}^{N-1} e^{\frac{j2\pi(n-m)k}{N}} \mathbb{E}(\tilde{\mathbf{h}}_k^T \tilde{\mathbf{h}}_k^*) \\ &= \frac{1}{N} \sum_{k=0}^{N-1} e^{\frac{j2\pi(n-m)k}{N}} \sum_{l=0}^{L-1} \mathbb{E}(|\tilde{h}[l, k]|^2). \end{aligned} \quad (\text{A.12})$$

Following the derivation of MSE in [17], it is easy to obtain that

$$\mathbb{E}(|\tilde{h}[l, k]|^2) = L \left(\frac{\sigma_w^2}{N\sigma_d^2} + \frac{2N^2 - 8kN + 8k^2}{N^3} \cdot \frac{\sigma_w^2}{\sigma_t^2} \right), \quad (\text{A.13})$$

and we therefore have

$$\begin{aligned} \mathbb{E}(\tilde{\mathbf{H}}_m^T \tilde{\mathbf{H}}_n^*) &= \frac{L}{N} \sum_{k=0}^{N-1} e^{\frac{j2\pi(n-m)k}{N}} \left(\frac{\sigma_w^2}{N\sigma_d^2} + \frac{2N^2 - 8kN + 8k^2}{N^3} \cdot \frac{\sigma_w^2}{\sigma_t^2} \right). \end{aligned} \quad (\text{A.14})$$

Further derivation of (A.14) depends on the relation between m and n .

A. $m \neq n$

For this situation, (A.14) can be rewritten as

$$\begin{aligned} \mathbb{E}(\tilde{\mathbf{H}}_m^T \tilde{\mathbf{H}}_n^*) &= \frac{L}{N} \left[\sum_{k=0}^{N-1} e^{\frac{j2\pi(n-m)k}{N}} \left(\frac{\sigma_w^2}{N\sigma_d^2} + \frac{2\sigma_w^2}{N\sigma_t^2} \right) \right] \\ &+ \frac{L}{N} \left[\sum_{k=0}^{N-1} e^{\frac{j2\pi(n-m)k}{N}} \left(\frac{-8k\sigma_w^2}{N^2\sigma_t^2} \right) \right] + \frac{L}{N} \left[\sum_{k=0}^{N-1} e^{\frac{j2\pi(n-m)k}{N}} \left(\frac{8k^2\sigma_w^2}{N^3\sigma_t^2} \right) \right]. \end{aligned} \quad (\text{A.15})$$

Since $m \neq n$, the first summarization in (A.15) is equal to zero. Using the relation [24]

$$\sum_{k=0}^{N-1} kq^k = \frac{-(N-1)q^N}{1-q} + \frac{q(1-q^{N-1})}{(1-q)^2}, \quad (\text{A.16})$$

we find that the numerator of the second summarization is a linear function of N . Considering that the denominator for the second summarization is a quadratic function of N , the second term is also zero when N is large enough. Similarly, bearing in mind the following relation [24],

$$\begin{aligned} \sum_{k=0}^{N-1} k^2 q^k &= \frac{-(N-1)^2 q^{N+2} + (2N^2 - 2N - 1)q^{N+1} - N^2 q^N + q^2 + q}{(1-q)^3} \end{aligned} \quad (\text{A.17})$$

and the fact that the denominator for the third summarization term is a cubic function of N , the third summarization term is equal to zero as well for large enough N . Since all the three terms are zeros, we can conclude that $\mathbb{E}(\tilde{\mathbf{H}}_m^T \tilde{\mathbf{H}}_n^*) = 0$ given $m \neq n$.

B. $m = n$

Given $m = n$, (A.14) can be rewritten as

$$\mathbb{E}(\tilde{\mathbf{H}}_m^T \tilde{\mathbf{H}}_m^*) = \frac{L}{N} \sum_{k=0}^{N-1} \left(\frac{\sigma_w^2}{N\sigma_d^2} + \frac{2N^2 - 8kN + 8k^2}{N^3} \cdot \frac{\sigma_w^2}{\sigma_t^2} \right) \quad (\text{A.18})$$

which, after some algebra, can be simplified as

$$\mathbb{E}(\tilde{\mathbf{H}}_m^T \tilde{\mathbf{H}}_m^*) = \frac{L}{N} \left(\frac{\sigma_w^2}{\sigma_d^2} + \frac{2\sigma_w^2}{3\sigma_t^2} \right) \quad (\text{A.19})$$

for large enough N .

In view of the two situations above, we can finally obtain (20).

Appendix B

As shown in [17], it is easy to derive that the estimates of v_l and h_l are both unbiased for large enough N . Therefore, we can also ascertain that the estimations of $h[l, k]$ are unbiased as well; that is,

$$\mathbb{E}(\tilde{h}[l, k]) = 0. \quad (\text{B.1})$$

Bearing in mind (5), the following relation can be obtained

$$\begin{aligned} \mathbb{E}\{[\tilde{\mathbf{H}}]_{(m,n)}\} &= \frac{1}{N} \sum_{l=0}^{L-1} \sum_{k=0}^{N-1} \mathbb{E}(\tilde{h}[l, k]) e^{\frac{j2\pi(n-m)k}{N}} e^{-\frac{j2\pi nl}{N}} \\ &= 0 \end{aligned} \quad (\text{B.2})$$

which means the estimations of \mathbf{H} are also unbiased; that is,

$$\mathbb{E}(\tilde{\mathbf{H}}) = \mathbf{0}, \quad (\text{B.3})$$

where $\mathbf{0}$ is an $N \times N$ zero matrix. In view of the conclusion

above, we can obtain

$$\begin{aligned} E(\hat{\mathbf{H}}\hat{\mathbf{H}}^H) &= E[(\mathbf{H} + \tilde{\mathbf{H}})(\mathbf{H} + \tilde{\mathbf{H}})^H] \\ &= E(\mathbf{H}\mathbf{H}^H) + E(\tilde{\mathbf{H}}\tilde{\mathbf{H}}^H). \end{aligned} \quad (\text{B.4})$$

In Appendix A, we have derived the expression of $E(\tilde{\mathbf{H}}\tilde{\mathbf{H}}^H)$. Recalling the MSE representation in [17] for large enough M , (20) can be rewritten as

$$E(\tilde{\mathbf{H}}\tilde{\mathbf{H}}^H) = \frac{1}{N} \overline{\text{MSE}} \cdot \mathbf{I}, \quad (\text{B.5})$$

where $\overline{\text{MSE}}$ denotes the average MSE. Since the definition of $\overline{\text{MSE}}$ is given as

$$\overline{\text{MSE}} = \sum_{l=0}^{L-1} \sum_{k=0}^{N-1} E(|\tilde{h}[l, k]|^2) \quad (\text{B.6})$$

for large N , (B.5) can be further rewritten as

$$E(\tilde{\mathbf{H}}\tilde{\mathbf{H}}^H) = \frac{1}{N} \sum_{l=0}^{L-1} \sum_{k=0}^{N-1} E(|\tilde{h}[l, k]|^2) \cdot \mathbf{I}. \quad (\text{B.7})$$

Recalling the similarity between the definition of \mathbf{H} and $\tilde{\mathbf{H}}$, it is easy to derive that

$$\begin{aligned} E(\mathbf{H}\mathbf{H}^H) &= \frac{1}{N} \sum_{l=0}^{L-1} \sum_{k=0}^{N-1} E(|h[l, k]|^2) \cdot \mathbf{I} \\ &= L\sigma_h^2 \cdot \mathbf{I}, \end{aligned} \quad (\text{B.8})$$

where σ_h^2 is defined in (36).

Since $E(\mathbf{H}\mathbf{H}^H)$ and $E(\tilde{\mathbf{H}}\tilde{\mathbf{H}}^H)$ are both diagonal matrices, we have

$$E(\hat{\mathbf{H}}\hat{\mathbf{H}}^H) = \left[L\sigma_h^2 + \frac{L}{N} \left(\frac{\sigma_w^2}{\sigma_d^2} + \frac{2\sigma_w^2}{3\sigma_t^2} \right) \right] \cdot \mathbf{I}. \quad (\text{B.9})$$

For (B.9), it is easy to observe that all the eigenvalues of $E(\hat{\mathbf{H}}\hat{\mathbf{H}}^H)$ are equal; thus, we have

$$\bar{\lambda}_{\max} = L\sigma_h^2 + \frac{L}{N} \left(\frac{\sigma_w^2}{\sigma_d^2} + \frac{2\sigma_w^2}{3\sigma_t^2} \right). \quad (\text{B.10})$$

References

- [1] J.A.C. Bingham, "Multicarrier Modulation for Data Transmission: An Idea Whose Time Has Come," *IEEE Commun. Mag.*, vol. 28, no. 5, May 1990, pp. 5–14.
- [2] T. May, H. Rohling, and V. Engels, "Performance Analysis of Viterbi Decoding for 64-DAPSK and 64-QAM Modulated OFDM Signals," *IEEE Trans. Commun.*, vol. 46, no. 2, Feb. 1998, pp. 182–190.
- [3] Y. Li, "Pilot-Symbol-Aided Channel Estimation for OFDM in Wireless Systems," *IEEE Trans. Veh. Technol.*, vol. 49, no. 4, July 2000, pp. 1207–1215.
- [4] M.J.F.-G. Garcia, S. Zazo, and J.M. Paez-Borrillo, "Pilot Patterns for Channel Estimation in OFDM," *Electron. Lett.*, vol. 36, no. 12, June 2000, pp. 1049–1050.
- [5] J.W. Choi and Y.-H. Lee, "Optimum Pilot Pattern for Channel Estimation in OFDM Systems," *IEEE Trans. Wireless Commun.*, vol. 4, no. 5, Sept. 2005, pp. 2083–2088.
- [6] K.F. Lee and D.B. Williams, "Pilot-Symbol-Assisted Channel Estimation for Space-Time Coded OFDM Systems," *EURASIP J. Adv. Signal Process.*, vol. 2002, no. 1, May 2002, pp. 507–516.
- [7] S. Coleri et al., "Channel Estimation Techniques Based on Pilot Arrangement in OFDM Systems," *IEEE Trans. Broadcast.*, vol. 48, no. 3, Sept. 2002, pp. 223–229.
- [8] S. Ohno and G.B. Giannakis, "Capacity Maximizing MMSE-Optimal Pilots for Wireless OFDM over Frequency-Selective Block Rayleigh-Fading Channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 9, Sept. 2004, pp. 2138–2145.
- [9] M. Visintin, "Karhunen-Loeve Expansion of a Fast Rayleigh Fading Process," *Electron. Lett.*, vol. 32, no. 18, Aug. 1996, pp. 1712–1713.
- [10] T. Zemen and C.F. Mecklenbrauker, "Time-Variant Channel Estimation Using Discrete Prolate Spheroidal Sequences," *IEEE Trans. Signal Process.*, vol. 53, no. 9, Sept. 2005, pp. 3597–3607.
- [11] Z. Tang et al., "Pilot-Assisted Time-Varying Channel Estimation for OFDM Systems," *IEEE Trans. Signal Process.*, vol. 55, no. 5, May 2007, pp. 2226–2238.
- [12] K.D. Teo and S. Ohno, "Optimal MMSE Finite Parameter Model for Doubly-Selective Channels," *IEEE Global Telecommun.*, St. Louis, MO, USA, Nov. 28–Dec. 2, 2005, pp. 3503–3507.
- [13] S. Haykin, "Adaptive Filter Theory," Englewood Cliffs, NJ, USA: Prentice-Hall, 1996, pp. 610–620.
- [14] F. Hlawatsch and G. Matz, "Wireless Communication over Rapidly Time-Varying Channels," Burlington, VT, USA: Academic Press of Elsevier, 2011, pp. 40–48.
- [15] Z. Du et al., "Maximum Likelihood Based Channel Estimation for Macrocellular OFDM Uplinks in Dispersive Time-Varying Channels," *IEEE Trans. Wireless Commun.*, vol. 10, no. 1, Jan. 2011, pp. 176–187.
- [16] Y. Liu et al., "Channel Estimation for Macrocellular OFDM Uplinks in Time-Varying Channels," *IEEE Trans. Veh. Technol.*, vol. 61, no. 4, May 2012, pp. 1709–1718.
- [17] R. Yao et al., "On Channel Estimation for OFDM Uplinks in Time-Varying Channels," to appear in *IET Commun.*, 2015.
- [18] H. Huang, "Spatial Channel Model for Multiple Input Multiple Output (MIMO) Simulations," 3GPP TSG RAN, Edinburgh, UK, Tech. Rep. 3GPP TR 25.996(12.0.0)-12, Sept. 2014.
- [19] A. Gorokhov and J.-P. Linnartz, "Robust OFDM Receivers for Dispersive Time-Varying Channels: Equalization and Channel Acquisition," *IEEE Trans. Commun.*, vol. 52, no. 4, Apr. 2004, pp. 572–583.

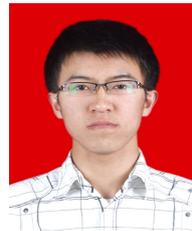
- [20] A. Scaglione, S. Barbarossa, and G.B. Giannakis, "Self-Recovering Equalization of Time-Selective Fading Channels Using Redundant Filterbank Procedures," in *Proc. DSP Workshop*, Bryce Canyon, UT, USA, Sept. 1998.
- [21] A. Goldsmith, "*Wireless Communications*," Cambridge, UK: Cambridge University Press, 2005, pp. 65, 303–308, 358–366.
- [22] D. Tse and P. Viswanath, "*Fundamentals of Wireless Communication*," Cambridge, UK: Cambridge University Press, 2005, pp. 332–382.
- [23] T. Yoo and A. Goldsmith, "Capacity and Power Allocation for Fading MIMO Channels with Channel Estimation Error," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, May 2006, pp. 2203–2214.
- [24] A. Jeffrey and D. Zwillinger, "*Table of Integrals, Series, and Products*," 7th Edition, Amsterdam, Netherlands: Elsevier Academic Press, 2007, pp. 1–23.



Rugui Yao received his BS, MS, and PhD degrees in communications from the School of Electronics and Information, Northwestern Polytechnical University (NPU), Xi'an, China, in 2002, 2005, and 2007, respectively. He worked as a post-doctoral fellow at NPU from 2007 to 2009. In 2010, he joined NPU as an associate professor. In 2013, he joined the ITP Lab of Georgia Tech, Atlanta, GA, USA, as a visiting scholar. He has worked in the area of cognitive radio networks, channel coding, OFDM transmission, and spread spectrum systems.



Yinsheng Liu received his BS degree in communication and information systems from North China Electric Power University, Baoding, China and his MS degree in communication and information systems from Beijing Jiaotong University, Beijing, China, in 2007 and 2009, respectively. He is currently pursuing his PhD degree in traffic information engineering and control at Beijing Jiaotong University. His research interests include wireless communication systems and digital signal processing.



Geng Li received his BS degree in communication engineering from the School of Electronics and Information, Northwestern Polytechnical University (NPU), Xi'an, China, in 2012 and is currently pursuing his MS degree in electronic and communication engineering at NPU. His research interests include wireless communications and anti-jamming techniques.



Juan Xu received her BS, MS, and PhD degrees all in computer science from the School of Computer Science, Northwestern Polytechnical University, Xi'an, China, in 2002, 2005, and 2011 respectively. From 2011, she joined the School of Electronic and Control Engineering, Xi'an, China. Her research interests include channel coding, OFDM transmission, and spread spectrum systems.